Estimating the Interest Rate Term Structure of Corporate Debt With a Semiparametric Penalized Spline Model

Robert Jarrow, David Ruppert, and Yan Yu

1. INTRODUCTION

This article contains a case study in statistical finance as well as methodological questions of broader interest. We suggest a new method of choosing the smoothing parameter when estimating the derivative of a function using a spline. In this article, we discuss mathematical as well as financial derivatives. The meaning should be clear from the context.

The prices of bonds determine an implied interest rate. Consider a zero-coupon bond paying no interest or principal until maturity, then paying a fixed amount called the par value. Suppose that $P(0, t)$ is the current (time 0) price, as a fraction of the par value, of a zero-coupon bond maturing in $t$ years. This price is consistent with a variable interest rate $f(0, t)$, called the forward rate, such that

$$P(0, t) = \exp\left(-\int_0^t f(0, s) \, ds \right).$$

(1)

The financial significance of $f(0, t)$ is that it is the rate one can lock in today for future borrowing or lending at time $t$. Figure 1(a) is a plot of $-\log(P)$ versus $t$ for typical price data for a zero-coupon bond. There are maturities from 0 to 30 years, spaced nearly quarterly. The rough linear increase of $-\log(P)$ appears consistent with $f(0, s)$ in (1) being nearly constant, but deviations from a constant rate are difficult to detect with this plot. One can also look at Figure 1(b). The "empirical forward" rate in that figure is $\Delta_l - (\log(P))/\Delta_l$, where $\Delta$ is the differencing operator. The empirical bias bandwidth selection (EBBS) and generalized cross-validation (GCV) estimates of the forward rate in Figure 1(b) are not obtained from fitting a model for the forward rate to the empirical forward rates but rather by fitting model (1) to the log-price data in Figure 1(a) as explained later. A key point is that the difference quotients exhibit both random variation and systematic deviation from a constant rate. The errors can be attributed, among other things, to steadiness of the price data due. The observed bond prices are either from quotes or from previous transactions. As such, these prices may occur at different times or be for different quantities.

The dependence of $f(0, t)$ on time to maturity $t$ is called the term structure. A more general description of the term structure is the evolution of interest rate function $f(s, s + t)$ at time $s$ over $t$ periods to maturity. As a function of $s, f(s, s + t)$ exhibits erratic random behavior, usually modeled as a Brownian motion rather than a smooth function. For this reason, $f(s, s + t)$ is estimated separately for each value of $s$. Therefore, current time $s$ will be fixed at 0, so $f(0, t)$ is denoted by $f(t)$ hereafter.

The term structure can only be inferred from observable bond prices. Although the literature studying the estimation of Treasury term structure is voluminous (see McCulloch 1971, 1975; Vasichek and Zorns 1978; Shea 1985; Chambers, Carleton, and Waldman 1984; Adams and Van Deventer 1994; Fisher, Nychka, and Zervos 1995), the literature studying corporate term structure estimation is much smaller (see Schwartz 1998 and references therein). The problem is that, for any individual corporation, there are bond prices at only a few maturities so determination of $f(t)$ for all $t$ is challenging. This appears to be the first article to estimate the term structure for bonds of an individual corporation.

There are many reasons estimation of $f$ is of interest. Suppose one were offering to buy or to sell a bond of a maturity not traded recently. Estimation of $f$ allows one to interpolate prices from other maturities. There are also more complex and interesting applications of the term structure. Corporate bonds are a classical example of a financial instrument bearing credit risk, the risk that an agent fails to fulfill contractual obligations.
Increased trading in instruments subject to credit risk has led to the creation of credit derivatives, instruments that partially or fully offset the credit risk of a deal. Given the recent explosive growth in the market for credit derivatives (see Risk Magazine 2002) and the regulatory-induced need to account for credit risk in the determination of equity capital (net worth of a business raised from owners), for example, Jarrow and Turnbull (2000), the estimation of corporate term structures has become of paramount interest. To put this in perspective, the size of the credit derivatives market in 2001 (as measured in notional amounts outstanding) was estimated to be 835.5 billion dollars.

The most traded credit derivatives include default swaps, credit-spread options, credit-linked notes, and collateralized default obligations (CDO's). For example, a credit call (put) option gives its owner the right to buy (sell) a credit-risky asset at a predetermined price, regardless of credit events that may occur before expiration of the option. A full treatment of credit derivatives can be found in Bielecki and Rutkowski (2002). The primary inputs to pricing models for these credit derivatives are the corporate term structures (see Jarrow and Turnbull 1995; Duffie and Singleton 1999; Bielecki and Rutkowski 2002). These term structures can also be used to infer the market’s assessment of credit quality for related uses in risk management procedures (see Jarrow 2001). Credit quality assessment is essential for value at risk (VaR) computations, bond portfolio management, corporate loan considerations, and even FDIC insurance premium calculations (see FDIC 2000).

In the estimation of the Treasury term structure, hundreds of bond prices are normally available on any given month, but for corporate term structures only a handful usually exist. This problem is observed in the Fixed Income database (Warga 1995). Consequently, corporate bonds require special estimation procedures.

Fisher, Nychka, and Zervos’s (1995) (F–N–Z) penalized spline model is nonparametric and, as such, requires numerous bond price observations. The F–N–Z model applies to Treasury bonds where prices at many maturities are available on any date, but it is problematic when applied directly to corporate debt. We generalize the F–N–Z model to corporate debt by modeling the corporate term structure as a Treasury term structure plus a parametric spread. The spread is the extra interest investors demand to buy risky and less liquid corporate bonds instead of Treasury bonds. For the Treasury term structure, we use F–N–Z’s nonparametric model. We find that a credit spread that is constant in time, thus requiring only a single parameter, fits our data well. In other situations, a spread that is linear in time might be used.

We extend F–N–Z’s work by (a) providing a comparison of generalized cross-validation (GCV), Ruppert’s (1997) EBBS method, and Ellner and Seif’s (2002) residual spatial autocorrelation (RSA) method for choosing penalty parameters, (b) deriving asymptotic sampling distributions for the term structure estimates, which enable us to compute confidence bands for the term structure estimates.

The term structure of interest rates can be identified by any one of four functions: the discount function, the yield curve, the forward rate curve, or the definite integral of the forward rate. Each one of these determines the other three. The forward curve has already been discussed. The discount function, $D(t)$, gives the price of a zero-coupon bond that pays one dollar at maturity time $t$, so that $D(t) = P(t)$ is given by (1). The yield curve, $y(t)$, is the average of $f(s)$ between 0 and $t$: $y(t) = \frac{1}{t} \int_0^t f(s) \, ds$. The definite integral of $f$ is $F(t) = ty(t)$. The relationships among these functions is

$$P(t) = D(t) = \exp[-F(t)] = \exp[-ty(t)] = \exp \left\{ - \int_0^t f(s) \, ds \right\}.$$  \hspace{1cm} (2)

Should one use a smoothing spline model for the forward rate $f$ or for some other function such as $D(t)$? F–N–Z considered spline modeling of $f$, $F$, and $D$ and concluded that modeling $f$ results in the most accurate estimation. If $D$ is modeled as a spline, then the model is linear in the spline coefficients, which is obviously attractive. However, there are advantages to
modeling $f$ itself as a spline. The constraint that a dollar paid today is worth a dollar, that is, that $D(0) = 1$, is then embedded in this model. In contrast, when fitting splines to $D$, the constraint $D(0) = 1$ must be imposed. Also, Shea (1985) noticed serious problems fitting splines to $D$, such as negative forward rates and instability at the long maturities. For these reasons, in this article, as in F–N–Z, $f$ will be modeled as a spline. However, differentiation of a spline produces another spline of lower degree so $f$ is a degree $p$ spline if and only if $F$ is a degree $p + 1$ spline. The distinction between whether $f$ or $F$ is modeled by a smoothing spline with the usual penalty on the second derivative really is a question of whether the roughness penalty is put on $f''$ or $F'' = f'$. 

Equation (2) holds only for zero-coupon bonds, but many bonds, including the AT&T bonds in our case study, have coupons. To price a coupon bond, we can view that bond as a portfolio of zero-coupon bonds, one for each payment. Payments can be priced by (2) and then summed. Let $P_1, \ldots, P_n$ denote the current (time 0) observed market prices of $n$ bonds from which the interest rate term structure is to be inferred. Bond $i$, $i = 1, \ldots, n$, has $z_i$ fixed payments $C_i(t_{i,j})$ due on dates $t_{i,j}$, $j = 1, \ldots, z_i$. The payment, $C_i(t_{i,j})$, consists of interest only for $j < z_i$ and principal and interest at maturity, $j = z_i$. The model price for the $i$th coupon bond is

$$
\hat{P}_i(\delta) = \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp[-t_{i,j} y(t_{i,j})] = \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp \left[ - \int_0^{t_{i,j}} f(s, \delta) \, ds \right],
$$

where $\delta$ is a vector of parameters in the model $f(s, \delta)$ for $f(s)$. We adopt a penalized splines (P splines) approach to the forward rate estimation. P splines are a generalization of smoothing splines that allow more general placement of knots and penalties. A relatively large number, $K$, of knots is used, but typically far less than for a smoothing spline; for example, a P spline may use $K = 20$ for $n = 200$. Once the number of knots is selected, the knots are located at equally spaced points as in Eilers and Marx (1996) or, as in Ruppert and Carroll (2000) and in Section 7, at equally spaced quantiles of the independent variable. The roughness penalty prevents overfitting, the value of $K$ is not crucial, provided that more than a minimum value is used; see Ruppert (2002). One could use $K = n$ as for smoothing splines, but doing this only increases the computational burden. F–N–Z called their estimators smoothing splines, but they also used far less than $n$ knots so we consider the F–N–Z estimators also to be P splines, not smoothing splines as the latter are defined in the literature, see, for example, Wahba (1990).

P splines, like their special case of smoothing splines, minimize the sum of a goodness-of-fit statistic plus a roughness penalty. We model the spline as $f(s, \delta) = \delta' B(s)$, where $B(s)$ is a vector of spline basis functions and $\delta$ is a vector of spline coefficients. Therefore, $F(t) = \int_0^t f(s) \, ds = \int_0^t (t - s) \delta' B(s) \, ds$. The roughness penalty is $\lambda \delta' G \delta$, where $\lambda > 0$ is a smoothing parameter and $G$ is a symmetric, positive semidefinite matrix. Possible choices of $G$ are discussed in Section 3. If $B(s)$ are splines of degree $p$, then $\delta$ determines the jumps in the $p$th derivative of $f$ or the $(p + 1)$st derivative of $F$ and $\lambda \delta' G \delta$ penalizes those jumps.

Proper selection of $\lambda$ to control the trade-off between goodness of fit and smoothness is crucial but complicated by three difficulties. The first, that GCV uses the trace of the smoother matrix defined only for linear smoothers, is solved by F–N–Z’s approximation based on a Taylor linearization. Another possible solution to this problem, one that we will study, is to fit $F(t) = - \log(P)$, which is a linear smoothing problem for zero-coupon bonds.

A second problem is that the choice of the smoothing parameter depends on the function estimated. We are estimating $f(t) = (d/dt) \left[ - \log(D(t)) \right]$. but because least squares compares $D(t)$ to prices or $F(t)$ to $- \log(D(t))$, GCV will choose the $\lambda$ best for estimating $D$ or $F$, not $f$. It is well known that the amount of smoothing that is optimal for estimation of a function is not the same as for estimating a derivative of a function. Asymptotics, for example, for local polynomial regression (Ruppert and Wand 1994), show that the amount of smoothing optimal for a first derivative decreases to 0 at a slower rate than for the function itself. The empirical evidence is that GCV tends to undersmooth the estimate of $f$. This undersmoothing is seen clearly in figures 5 and 7 of F–N–Z and also in Figure 1(b) of this article. Estimates of $f(t)$ often rise or fall rapidly as $t$ varies from 15 or 20. It is difficult to believe, for example, that the rate for three-month borrowing 23 years forward is 4% whereas the rate of three-month borrowing 21 or 25 years forward is over 6%. However, the estimates often show such behavior because of undersmoothing. Practitioners prefer a smooth forward curve for sound reasons, and the title of Adams and Van Deventer’s (1994) article emphasizes this preference. We address this undersmoothing problem by a modification of the EBBS method of Ruppert (1997), which minimizes an estimate of the mean square error of $f$.

A third problem is that GCV and related methods such as cross-validation (CV) assume independent errors. This assumption is suspect in our case. Some bonds trade at premium because of special liquidity or other advantages (Tuckman 2002). These bonds have lower yields (higher prices) and bonds of nearby maturities are close substitutes and also trade at a premium. Such premiums are not part of the term structure because they do not apply to other types of bonds. Thus, these premiums result in a cluster of correlated and more variable errors. Because of possible correlation, an alternative method of smoothing parameter selection (Eilers and Seif u2002) based on RQA is considered. However, in the case study we find that RQA and standard GCV undersmooth whereas EBBS works better. We tried correcting EBBS for autocorrelation, but found that this correction has little effect on the amount of smoothing chosen by EBBS.

The F–N–Z method of GCV introduces an additional parameter $\theta$ to control the amount of smoothing, as will be explained soon. However, the usual justification for using GCV is that it approximates CV (cross-validation), but this is only true when $\theta = 1$, which is the standard choice. F–N–Z provided no theoretical justification for introducing $\theta$ or using $\theta \neq 1$, but we believe EBBS explains why using $\theta = 2$ works better than using $\theta = 1$. F–N–Z’s GCV with $\theta = 2$ chooses a value of the smoothing parameter that is closer to the EBBS choice whereas.
standard GCV with $\theta = 1$ chooses less smoothing. Thus, using $\theta = 2$, as F–N–Z suggested, comes closer to the minimizing mean square error of the forward rate than using standard GCV. However, even F–N–Z’s version of GCV smooths less than EBBS.

We present a case study of U.S. Treasury STRIPS and AT&T bonds on December 1995. We then repeat the analysis 21 times independently, once for each of the earlier months over the period from April 1994 to December 1995. A Treasury STRIPS (Separate Trading of Registered Interest and Principal of Securities) is a synthetic zero-coupon bond constructed from Treasury bonds and issued by the Federal Reserve (Jarrow 2002). The AT&T bonds bear coupons. The data are from the University of Houston Fixed Income database (Warga 1995). There are two ways to estimate corporate term structure. The one-step method simultaneously estimates the Treasury term structure and the credit spread for a single corporation by minimizing the penalized sum of squares between the model prices and the observed market prices of the Treasury bonds and corporate bonds. In the two-step procedure, first one estimates the nonparametric Treasury term structure and then, with that fixed, estimates the credit spread by minimizing the nonpenalized sum of squares between the market and model prices of the corporate bonds. The two-step procedure is motivated by the application at hand. Although only one Treasury term structure exists, there are thousands of different corporate term structures, one for each company issuing debt. It makes sense to estimate the Treasury term structure only once, so we recommend and use the two-step procedure.

Section 2 describes the Fixed Income database. Section 3 introduces P splines and presents a spline model for Treasury bonds. Section 4 discusses the GCV, RSA, and EBBS criteria for selecting the penalty parameter. Section 5 describes the two-step estimation procedure. Asymptotics, confidence bands, and tests about the credit spread model are presented in Section 6. The case study is presented in Section 7.

2. DATA

The University of Houston Fixed Income database includes over 28,000 instruments and covers virtually every firm that has outstanding publicly traded nonconvertible debt with principal value of at least one million dollars. Information on individual bonds that make up the Lehman Brothers Bond Indices are reported, including month-end flat prices, accrued interest, coupon, yields, current date, issuance date, maturity date, S&P and Moody’s ratings, and option-like features.

The data for our case study consist of all U.S. Treasury STRIPS (coupon and principal STRIPS, that is, zero-coupon bonds that are synthesized from the coupon and principal payments of Treasury bonds) and all AT&T bonds. Market prices are available for five AT&T bonds on December 31, 1995. All have semiannual coupons with different maturities and with no embedded option features, for example, the right to prepay, for which our price model does not apply. Each price is obtained from the quoted flat price plus accrued interest. See Table 1.

Issue and maturity are given in year–month–day format. We need the time to maturity and the coupon payment times, $t_{i,k}$, on the same scale. The MATLAB finance toolbox can easily handle date conversions using, for instance, the functions \texttt{days365(·)} and \texttt{days360(·)}, for dates based on 365 or 360 days a year; 30-day months or 360 days per year is a convention used for some types of bonds, but not those in our case study. The coupon payment time can then be calculated by the function \texttt{cfdates(·)}. These calculations can also be easily implemented if the day counts need to exclude holidays and weekends. We use MATLAB functions \texttt{days365(·)} and \texttt{cfdates(·)} based on conventional actual/365 day count.

Table 2 lists the summary statistics for the numbers of U.S. Treasury STRIPS and AT&T bonds available over the 21-month period of April 1994 to December 1995 and demonstrates that far fewer AT&T bonds are available than U.S. Treasury STRIPS.

3. A SPLINE MODEL FOR THE TERM STRUCTURE OF TREASURY BONDS

The Treasury forward rate curve, denoted $f_{Tr}$, will be approximated by a spline $f_{Tr}(t) = \delta B(t)$. Here $B(t)$ is a vector of spline basis functions, for example, (truncated) power basis functions or $B$ splines, and $\delta$ is the coefficient vector. We
will use the $p$th degree power basis functions, with $B(t) = (1, t, \ldots, t^p, (t - \kappa_1)^{p+1}, \ldots, (t - \kappa_k)^{p+1})'$ and $\delta = (\delta_0, \delta, \ldots, \delta_{p+k})'$, where $[\kappa_{k+1}] = \kappa_k$ are spline knots and $(t - \kappa_k)^{p+1}$ if $k \geq k$ and is 0 otherwise. The power basis is convenient because polynomial submodels can be defined by setting certain coefficients to 0. This basis can be poorly conditioned if there are many knots, but a penalty usually solves this problem. Moreover, we computed using an orthogonal basis and then transformed the results back to the power basis; see Ruppert (2002). The forward rate $f_{tr}(t)$ is estimated by minimizing

$$Q_{a,\lambda} (\delta) = \frac{1}{n} \sum_{i=1}^{n} \left[ h(P_i) - h(\hat{P}_i(\delta)) \right]^2 + \lambda \delta' G \delta,$$

where $h$ is some monotonic transformation, for example, the identity or log function. For zero-coupon bonds, an advantage of the log function is that $h(\hat{P}_i(\delta))$ is linear in the parameters so that minimization of (4) is linear ridge regression. For statistical efficiency, the best choice of $h$ is one that gives residuals appearing closest to being normally distributed with a constant variance, but a small amount of heteroscedasticity has little negative effect on efficiency (Carroll and Ruppert 1984). Both the log and the identity transformations are suitable on these grounds. We found little difference in the estimates of $f$ using the identity and log transformations. The crucial choice is the parameter $\lambda$, not the transformation. We will report results for $h = \log$ for the STRIPS data.

There are several sensible choices for $G$. One choice is given by Ruppert, Wand, and Carroll (2003), where the power basis functions are used and $G$ is a diagonal matrix with its last $K$ diagonal elements equal to 1 and all others 0. This $G$ penalizes jumps at the knots in the $p$th derivative of the spline. As $\lambda \to \infty$, the fit converges to a $p$th degree polynomial fit. This penalty can be viewed as a penalty on the $(p+1)$st derivative where that derivative is a generalized function. In our numerical work in Section 7.1, we will use this penalty with $p = 2$.

A second choice, the quadratic penalty on the $d$th derivative, $\int B_d'(t)^2 dt$, for $d \leq p$, uses $G_{ij} = \int B_j'(t)^2 B_i(t) dt$, where $B_j(t)$ is the $j$th element of $B(t)$. Estimates using this penalty with $d = p = 2$ are similar to our numerical results in Section 7.1. The choice $d = 2$, the usual choice for smoothing splines, penalizes any deviation from linearity. In their maximum smoothness approach, Adams and Van Deventer (1994) used $d = 2$. Frischling and Yamamura (1997) used $d = 1$, which penalizes deviations from a constant function. If a quadratic integral penalty is used, then as $\lambda \to \infty$ the estimated forward rate converges to the $(d - 1)$st degree polynomial fit. Like the choice of the transformation $h$, we found the choice of $G$ relatively unimportant. Different choices give similar estimates.

We can rewrite expression (4) in a more revealing form. Because splines are piecewise polynomials, it is easy to compute their integrals. In our model, $f_d'(t) = \delta' f'(t)$, $df(s) = \delta' f'(t) dt$. Note

$$B'(t) = \int_0^t B(s) ds$$

$$= \left( \begin{array}{c}
\int_0^t B_1(s) ds \\
\int_0^t B_2(s) ds \\
\vdots \\
\int_0^t B_{p+1}(s) ds \\
\int_0^t (t - \kappa_1)^{p+1} ds \\
\int_0^t (t - \kappa_2)^{p+1} ds \\
\vdots \\
\int_0^t (t - \kappa_k)^{p+1} ds \\
\end{array}\right).$$

We can simplify the notation by expressing the model price in terms of $B'$ so that (4) becomes

$$Q_{a,\lambda} (\delta) = \frac{1}{n} \sum_{i=1}^{n} \left[ P_i - \sum_{j=1}^{n} C_i(t_{i,j}) \exp(-\delta' B'(t_{i,j})) \right]^2 + \lambda \delta' G \delta.$$  (5)

4. SELECTION OF THE KNOTS AND SMOOTHING PARAMETER

4.1 Choosing the Knots

An advantage of our P-spline approach is that the knots can be chosen automatically; following Ruppert, Wand, and Carroll (2003) and Ruppert (2002), the knot $\kappa_1$ is the $k/(k+1)$th sample quantile of $\{t_{i,j} : i,j=1\}$. Ruppert (2002) presented a detailed study of the choice of $K$. We recommend that $K$ be sufficiently large, say 8 or more, to accommodate nonlinearity of $f_{tr}$, but a larger $K$ does not cause overfitting provided $\lambda$ is suitably chosen.

4.2 Generalized Cross-Validation

A smoother is linear if $P$, the vector of observed responses, and $\hat{P}$, the vector of fitted values, are related by $P = A(\lambda) P$ for some "smoother" matrix $A(\lambda)$ independent of $P$. GCV is an approximation to cross-validation (CV) where $\lambda$ is chosen by minimizing

$$GCV(\lambda) = \frac{n^{-1} \sum_{i=1}^{n} \left[ P_i - \hat{P}_i(\delta) \right]^2}{\{n - n^1 - \theta \tr(A(\lambda))\}^2},$$

where $\theta$ is the tuning parameter introduced by F–N–Z and the trace of $A(\lambda)$ is the effective degrees of freedom. GCV used in the smoothing and penalized spline literature does not use this tuning parameter, so $\theta = 1$ corresponds to standard GCV. F–N–Z introduced $\theta$ because they found that standard GCV undersmoothed. We feel that we can provide an explanation for this undersmoothing. GCV does provide smooth estimates of the model function $D(t)$ but not of the derivative $f(t)$. This is understandable because GCV is based on the fit to the price data. BBSS does attempt to estimate $f(t)$ optimally. GCV with $\theta = 2$ mimics EBSS, which at least partially explains why the use of this extra tuning parameter is beneficial.

GCV applies to linear smoothers, but our smoother is linear only for zero-coupon bonds and $h = \log$. F–N–Z Taylor-expand the model about $\hat{\delta}$ to linearize. Let the model price of the $i$th bond be $m_i(\delta) := \hat{P}_i(\delta) = \sum_{j=1}^{n} C_i(t_{i,j}) \exp(-\delta' B'(t_{i,j})).$ Also, $m(\hat{\delta})(\delta) = (m_1(\hat{\delta})(\delta), \ldots, m_n(\hat{\delta})(\delta))^t$, where $m_i(\hat{\delta})(\delta) = \delta m_i(\hat{\delta})(\delta) = \sum_{j=1}^{n} C_i(t_{i,j}) \times \exp(-\delta' B'(t_{i,j})) B'(t_{i,j}).$ Define

$$\Sigma_n = n^{-1} \{ (m(\hat{\delta})'(\delta))' m(\hat{\delta})'(\delta) \}. \quad (7)$$

The approximate smoothing matrix is $A(\lambda) = m(\hat{\delta})'(\delta) [n(\Sigma_n + \lambda G)]^{-1} m(\hat{\delta})'(\delta)].$

4.3 Residual Spatial Autocorrelation

The RSA method of Ellner and Seifu (2002) applies Moran’s index of spatial autocorrelation $I$ to the residuals to choose the $\lambda$ giving the least deviation of $I$ from its expectation under random permutations. We refer the reader to Ellner and Seifu (2002) for details.
4.4 EBBS

EBBS developed by Ruppert (1997) for choosing the bandwidth for local regression can be extended to other smoothing parameters. Here we modify EBBS for use with P splines. EBBS models the bias of the fitted values as a function of the smoothing parameter. The variance of the fitted values can be estimated by an asymptotic formula or, using $h = \log \text{and zero-coupon bonds}$, by an exact result. To estimate the MSE, the estimated bias is squared and added to the estimated variance. When applied to $f_{\hat{r}}$, we have $\text{MSE}(f_{\hat{r}}, t, \lambda)$, the estimated MSE of $f_{\hat{r}}$ at $t$ and $\lambda$. MSE$(f_{\hat{r}}, t, \lambda)$ can be averaged over maturities $i \in \{1, \ldots, n\}$, and then minimized over $\lambda$. The variance of $\hat{f}_{\hat{r}}(t, \lambda)$ is estimated by (8) and (9) in Sections 6.2 and 6.3.

EBBS estimates bias at any fixed $t$ by computing the fit at $t$ for a range of values of the smoothing parameter and then fitting a curve to model bias. Our implementation of EBBS for P splines uses the fact that, to the first order, the bias of a P spline at $\lambda$ is $\gamma(t)$, for some $\gamma(t)$ (Wand 1999). Let $f_{\hat{r}}(t, \lambda)$ be $f_{\hat{r}}$, depending on maturity and $\lambda$. Compute $\{f_{\hat{r}}(t, \lambda_t), \ell = 1, \ldots, L\}$, where $\lambda_1 \leq \cdots \leq \lambda_L$ is the grid of values of $\lambda$ used for selecting $\lambda$ by GCV. For this grid, in our example we used $L = 50$ values of $\lambda$ such that their logarithms to base 10 were equally spaced between $-7$ and 1, which gave degrees of freedom values between $\text{DF}(10) = 4.8$ and $\text{DF}(10^{-7}) = 28.9$ for a 40-knot cubic spline fit. From experience with other smoothing problems, we felt that the optimal degrees of freedom would be somewhere in this range. For other examples of similar size, we recommend experimentation with the grid of $\lambda$ values to get roughly 5 to 30 degrees of freedom. For any fixed $t$, fit a straight line to the data $\{(f_{\hat{r}}(t, \lambda_i), i = 1, \ldots, L\}$. Let the slope of the line be $\hat{\gamma}(t)$. Then the estimate of squared bias at $t$ and $\lambda_t$ is $(\hat{\gamma}(t) \lambda_t)^2$.

5. THE TWO-STEP ESTIMATION PROCEDURE

Because there is only one Treasury curve but many corporate bond types (and credit spreads), we recommend the two-step procedure, now discussed in more detail.

Step 1 (Nonparametric P-spline fitting of a forward rate to U.S. Treasury bonds). The Treasury forward rate curves $f_{\hat{r}}$ or, equivalently, $\hat{\delta}$ is estimated by minimizing $Q_{n, \lambda}(\hat{\delta})$ in (5), and $\lambda$ is chosen by GCV, RSA, or EBBS as discussed in Section 4. Then $f_{\hat{r}}(t) = \hat{\delta} \mathbf{B}(t)$, where $\hat{\delta}$ are the estimated spline coefficients.

Step 2 (Parametric estimation to obtain the forward rate curve for a Corporation's bonds). The forward rate of a corporation's bonds is modeled as $f_{\hat{r}}(t) = f_{\hat{r}}(t) + \text{polynomial spread}$, with $f_{\hat{r}} = \hat{f}_{\hat{r}}$ from the first step. For example, if the credit spread is a constant, then we can write

$$f_{\hat{r}}(t) = \hat{f}_{\hat{r}}(t) + \alpha = \hat{\delta} \mathbf{B}(t),$$

where $\hat{\delta} = (\hat{\delta}_0 + \alpha, \hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_L)$. Consequently, the spread parameter $\alpha$ can be obtained by parametric nonlinear least squares estimation that minimizes the nonpenalized sum of squares difference between the observed market prices and the model prices of the corporate bonds.

We adopt polynomial spreads of low degree for several reasons. There are only five AT&T bond prices, so a simple parametric model is necessary. As can be seen in Table 1, the maturities of AT&T bonds are between 6 and 11.2 years, so estimation of the spread for $t > 11.2$ is extrapolation, with well-known dangers, and there is also relatively little information about the spread for $t < 6$. Using a simple parametric model of the credit spread will cause some bias, but this bias should be small. The estimated forward rate is between .05 and .07, but the estimated credit spread is an order of magnitude smaller, about .005. Therefore, the credit spread will be modeled by a constant term $\alpha$ (constant credit spread), by $\alpha + \beta t$ (linear credit spread), or by $\alpha + \beta t + \gamma t^2$ (quadratic credit spread). The spread parameters can be estimated by parametric nonlinear least squares with $\hat{\delta}$ fixed at $\hat{\delta}$ from step 1.

6. ASYMPTOTIC PROPERTIES AND INference

In this section, we develop asymptotic properties needed for inference and to justify the linearized GCV in Section 4.2. In the following, $n$ is the number of Treasury bond prices. We only study large-sample asymptotics for Treasury prices. The number of corporate bonds is usually so small that large-sample theory seems pointless.

Asymptotics could be developed with $K \to \infty$ as $n \to \infty$, but a fixed-$K$ asymptotic is most relevant for applications where large-sample theory provides approximate distributions. These approximations should be most accurate if $K$ is held at the value used in an application.

Because $K$ is fixed, consistency will mean convergence of $\hat{\delta}$ to $\delta_0$ defined as follows. Assume that the empirical distribution of $\{\tau, t\} = 1, \ldots, L$ converges weakly to some limiting distribution $F_X$. Consider the space of splines with knots equal to $F_X^{-1}(t/(K + 1))$, $t = 1, \ldots, K$. Then $\delta_0$ is defined as the coefficients of the spline that best approximate $f_{\hat{r}}$ in $L^2(F_X)$. Ruppert's (2002) results suggest that, for smooth $f_{\hat{r}}$, the best $L^2$ approximation is quite close to $f_{\hat{r}}$, and the bias due to approximating $f_{\hat{r}}$ by a spline is negligible compared to the standard deviation of $\hat{f}_{\hat{r}}$ and the bias of $\hat{f}_{\hat{r}}$ due to the penalty.

6.1 Large-Sample Asymptotics With $\lambda_n \to 0$

Denote $\lambda$ by $\lambda_n$. The variance of $\hat{\delta}$ goes to 0 as $n$ tends to $\infty$ whether $\lambda_n$ tends to 0. However, if $\lambda_n \to 0$ as $n \to \infty$, then the bias also tends to 0 and consistency can be established. The assumptions of the following two theorems are given in the Appendix. The proofs are similar to those in Yu and Ruppert (2002) and are omitted.

**Theorem 1.** Let $\{\hat{\delta}_{n, \lambda_n}\}$ be a sequence of penalized least squares estimators minimizing (5). Under Assumption A.1, if the smoothing parameter $\lambda_n$ is $o(1)$, then $\hat{\delta}_{n, \lambda_n}$ is a (strongly) consistent estimator of the true parameter $\delta_0$.

**Theorem 2.** Let $\{\hat{\delta}_{n, \lambda_n}\}$ be a sequence of penalized least squares estimators of (5). Under Assumptions A.1 and A.2, if the smoothing parameter $\lambda_n$ is $o(n^{-1/2})$, then $\sqrt{n}(\hat{\delta}_{n, \lambda_n} - \delta_0)$ $\xrightarrow{D}$ N(0, $\sigma^2\Omega^{-1}(\delta_0)$), where $\Omega(\delta_0) := \lim_n\Sigma_n$, is defined in (7).
6.2 Large-Sample Asymptotics With $\lambda$ Fixed and the Sandwich Formula

The asymptotic variance in Theorem 2 does not involve $\lambda$ because $\lambda$ goes to 0. In finite samples, this asymptotic variance will overestimate the variance of $\hat{\lambda}$, which is decreasing in $\lambda$, so for inference we give the asymptotic distribution of $\hat{\lambda}$ when $\lambda$ is fixed.

Using estimating equations, for example, in Carroll, Ruppert, and Stefanski (1995), we can derive the “sandwich formula” for the asymptotic variance matrix of $\hat{\lambda}(\lambda)$. From (5), $\hat{\lambda}(\lambda)$ is the solution to the estimating equation

$$0 = \frac{\partial}{\partial \delta} Q_{\beta\alpha}(\delta) = \sum_{i=1}^{n} \psi_i(\delta, \lambda, G),$$

where $\psi_i(\delta, \lambda, G) = -\{P_i - m_i(\delta) \gamma_i^{(1)}(\delta) + \lambda G\delta$. The sandwich formula for the asymptotic variance matrix of $\hat{\lambda}(\lambda)$ is $\text{Var}(\hat{\lambda}(\lambda)) = n^{-1} B_n^{-1} A_n B_n^{-1}$, where $A_n = n^{-1} \sum_{i=1}^{n} E[\psi_i(\delta, \lambda, G)] = \sigma^2 \Sigma_n$, and, with $\Sigma_n$ as in (7), $B_n = (\partial/\partial \delta') n^{-1} \sum_{i=1}^{n} E[\psi_i(\delta, \lambda, G)] = \Sigma_n + \lambda G$. Therefore

$$\text{Var}(\hat{\lambda}(\lambda)) = \sigma^2 \left[ \Sigma_n + \lambda G \right]^{-1} \Sigma_n \left[ \Sigma_n + \lambda G \right]^{-1}.$$

Note that, as $\lambda \to 0$, $\text{Var}(\hat{\lambda}(\lambda))$ converges to $n^{-1} \sigma^2 \Sigma_n^{-1}$ as given in Section 6.1.

6.3 Confidence Bands for $f_{Tr}$

Because the estimated Treasury forward rate at time $t$ for a Treasury STRIPS is $\hat{B}(t)$, a standard error for this forward rate is

$$\text{sd}(\hat{f}_{Tr}(t)) = \sqrt{\text{Var}(\hat{\lambda}(\lambda))} \left[ \hat{B}(t) \right],$$

where $\text{Var}(\hat{\lambda}(\lambda))$ is given by (8). By a delta method calculation, the standard error of the estimated discount function, $\hat{D}_{Tr}(t) = \exp(-\delta \hat{B}(t))$, is easily obtained. From these standard errors, pointwise confidence bands for $f_{Tr}(t)$ and $D_{Tr}(t)$ are obtained in the usual manner.

6.4 Bootstrap Inference About the Spread Parameters

Inference for nonlinear regression is often based on asymptotic theory but because the number of corporate bonds available is usually very small, large-sample theory seems dubious. Instead, we can handle the inference about the credit spread parameters via a parametric bootstrap. We will start first with the constant credit spread model. The parametric bootstrap algorithm is as follows:

1. Fix $f_{Tr}$, $\alpha$, and $\sigma$ as our estimates. These will be the “true” parameters in the bootstrap sampling. Use $f_{Tr}$ and $\alpha$ to generate the bond prices without observation error. These prices use the same maturities, coupon times, and coupon payments as in the original sample.
2. Repeat $N_b$ times (we use $N_b = 1000$):
   - Generate a bootstrap sample of corporate bond prices by adding $N(0, \sigma^2)$ errors to the error-free bond prices computed in the previous step.
   - Using the bootstrap sample, estimate $\hat{\delta}$ and compute the $F$ statistics for testing the hypothesis of interest, for example, for testing the null hypothesis of a constant credit spread versus a linear credit spread.

7. THE AT&T CASE STUDY

In this section, we return to our example of the term structure for AT&T bonds.

7.1 Estimation Results

We applied the two-step procedure separately to the end-of-the-month STRIPS and AT&T bond prices for each of the 21 months from April 1994 through December 1995. As an illustration, we will concentrate on prices on December 31, 1995. We used prices on this date as a test bed for our methods, especially for choosing the smoothing parameter. The negative logarithms of the STRIPS prices on this date are plotted in Figure 1. Because the STRIPS are zero coupon bonds, the negative log prices in Figure 1 follow the function $F(t) = ty(t)$. Therefore, we estimated $f(t)$ by fitting a cubic spline to the negative log prices and differentiating the estimate. Our main conclusions from estimating the Treasury forward rate are:

1. The GCV and RSA methods of smoothing parameter selection tend to undersmooth the estimate of $f_{Tr}$, with $f_{Tr}$ depending heavily on whether a small number of knots (say 8 or less) or a large number (20 or more) are used.
   - GCV or RSA could be used if the number of knots is chosen carefully, though this introduces subjectivity. If one uses a small number of knots, say 8, then GCV and RSA cannot overfit because the maximal degrees of freedom of a cubic spline is then 12, which provides a suitable amount of smoothing.
   - If GCV is used to select both $K$ and $DF(\lambda)$, then it will select $K$ and $D(\lambda)$ very large and produce a very rough fit.
   - GCV undersmooths when $\theta = 2$ is used as recommended by F–N–Z. Even $\theta = 3$ only smooths slightly more than $\theta = 1$ or 2. Figure 1 shows the forward rate estimates with $\theta = 1$ and 3. These estimates are similar to each other but dissimilar to the EBBS estimate. The curve for $\theta = 2$ is somewhere between those of $\theta = 1$ and 3.

2. In our case study, however, EBBS is stable with the fitted curve independent of the number of knots.
   - If EBBS and a cubic spline estimate for $F$ (quadratic for $f$) are used, then anywhere between 5 and 80 knots, and perhaps more, will work well, with $f_{Tr}$ depending very little on $K$.

3. The residuals are autocorrelated.
   - Autocorrelation does not affect EBBS much, though EBBS could be corrected for autocorrelation. Correcting for autocorrelation would increase the estimated variances of the fits but leave the estimated bias unchanged, so the forward rate would be smoothed somewhat more.
   - Autocorrelation is the main reason that GCV undersmooths.
   - Standard errors that assume independence are too small, but can be corrected.
The same conclusions are reached when using STRIPS prices on other dates. Note that the residual autocorrelation functions (ACFs) are nonstandard, because the indexing of the observations is by the time to maturity not by time of observation because the observations are all on the same date.

In our study, the GCV method uses a cubic power spline basis for $F(t)$ of 40 equally spaced quantile knots and quadratic penalty on the spline coefficients. F–N–Z uses a cubic B splines for $f(t)$. 1/3 data points as knots located at the data points, and the integral of the squared second derivative of the forward rate as the penalty. We ran the F–N–Z program in Mathematica and found as expected that it smooths the STRIPS data better with hyperparameter $\theta = 2$ than with $\theta = 1$. Our estimates using GCV are qualitatively similar to those from the F–N–Z program, and we believe that the somewhat different methods of estimation with GCV used here and in F–N–Z produce similar results. However, $\theta = 2$ still smooths less than EBBS and gives local maxima and minima in the estimated forward rate that are difficult to justify for financial reasons. We experimented with other bases such as B splines, with equally-spaced knots (instead of knots at equally spaced quantiles) and with different penalties, but our experience was that the fits were not changed much. The times to maturity are nearly equally spaced so it is not unexpected that using quantiles as knots was similar to using equally spaced knots. One difference between the F–N–Z study and ours is that F–N–Z used Treasury coupon bonds and we used STRIPS.

The EBBS residual sample ACF plot in Figure 2 where the autocorrelation at lag 1 is .87 shows that the residuals are highly autocorrelated when the amount of smoothing is large (as chosen by EBBS). The autocorrelation is less when the amount of smoothing is small (GCV’s selection). So the question is which estimate of autocorrelation to believe, EBBS’s or GCV’s. Because the GCV estimate is much rougher than practitioners in finance believe is realistic, we feel that there is serious correlation in the STRIPS prices and this is the reason that GCV undersmooths. Cross-validation and GCV are known to perform poorly in the presence of correlated noise (Hart 1991). We simulated STRIPS data with AR(1) noise with the same lag 1 autocorrelation as in the EBBS residuals and found that GCV greatly undersmoothed but EBBS did not. We also simulated STRIPS data with independent errors and found that GCV and EBBS gave similar estimates for most simulated datasets. However, about 10% of the time the GCV estimate does undersmooth substantially. So another potential problem with GCV is the well-known variability in the amount of smoothing it selects (Härdle, Hall, and Marron 1988).

If the errors are, in fact, positively correlated, then the variance of $\hat{\pi}$ is somewhat larger than indicated by the standard errors, which assume independence. To correct the standard errors, we could model the autocorrelation function, say with an ARMA model. Let $R_n$ then be the estimated correlation matrix of $P$. The corrected sandwich formula is

$$\text{Var}[\hat{\delta}(\lambda)] = \frac{\sigma^2}{n} [\{\Sigma_n + \lambda G\}^{-1}] C_n [\{\Sigma_n + \lambda G\}^{-1}]$$

where $C_n$ is the adjustment of $\Sigma_n$ in (7) for autocorrelation: $C_n = n^{-1}[[m^{(i)}(\delta)] R_n [m^{(i)}(\delta)]]$.

If the errors are correlated, then the premise behind RSA is false, so its ability to select a correct amount of smoothing is dubious. EBBS may also undersmooth because it underestimates variance, but during some Monte Carlo experimentation we
found that EBBS was much less susceptible to undersmoothing than GCV. Also, we modified the EBBS method by using (10) as the variance estimate. The corrected EBBS chose DF(λ) only somewhat smaller than the uncorrected EBBS and had little noticeable effect on the fit.

7.2 Modeling the Credit Spread Function

We can test hypotheses of economic interest about the credit spread via the bootstrap. A simple model that the credit spread is constant is tested by testing that β = 0 where the credit spread is α + βt. The F statistic for this hypothesis is 2.87 with a p value of .19 calculated by the proportion of the bootstrap F statistics that exceeds this value. If instead one tests a constant credit spread versus the alternative of a quadratic credit spread, then the p value is .37.

We compared the bootstrap distribution of the F statistics with the appropriate degrees of freedom (e.g., 1 and n_r - 2 for testing a constant credit spread versus linear spread). The true Type I error probabilities of nominal .1, .05, and .01 level F tests of a constant versus linear credit spread are .078, .033, and .008.

We also performed a power study for testing a constant spread versus a linear spread by using a linear spread with linear coefficient β varying. For β = 0, .001/12, .002/12, .005/12, .011/12, .012/12, and .015/12, the corresponding powers of the test are .05, .16, .21, .49, .86, .94, and .99 calculated by the proportion of F statistics that exceeds the 95th percentile of its bootstrap null distribution. The largest maturity is slightly more than 11 years, so β = .01/12 means the credit spread increases by about .01 from 0 until the longest maturity of the AT&T bonds.

The risk of default by AT&T immediately after time t = 0 seems negligible, and the credit spread at t = 0 should be due mostly to liquidity risk, not credit risk. [Marshall (2000) explained that "credit spread" is somewhat a misnomer because the spread is due to more factors than simply credit differences.] AT&T bonds are less liquid than Treasury bonds, so there is no guarantee that an AT&T bond holder could sell the bond immediately if that were necessary. If cash were needed quickly, the bond holder might need to sell at a discount to find an immediate buyer. This is liquidity risk. The intercept α of the credit spread can be interpreted as liquidity risk. In the constant spread model, the null hypothesis of no spread (α = 0) has an F statistic of 1261 and the p value is almost 0. Not surprisingly, there is extremely strong evidence that a spread exists.

Accepting the null hypothesis of a constant spread, of course, should not be interpreted as proving that the spread is constant. With only five corporate bonds, only substantial deviations from a constant spread could be detected. From the power study, we can see how large a deviation from a constant spread would need to be in order to be detected with high probability. A spread that increases linearly by 50 basis points (.5%) over 12 years has .5 probability of being detected by an F test. There are some arguments in favor of a constant spread. At the short end, there would be liquidity risk and the credit risk becoming smaller. At the middle and long end, there would still be liquidity risk and probably more credit risk. One might argue that the liquidity risk is less in the middle and long end because the bonds have a longer time to maturity and probably are more available for trading. To be honest, no one knows. So, it seems to us, as a first approximation, and given a minimal amount of bonds, a constant spread is a good first approximation, at least for AT&T bonds on December 1995. For other corporations or other dates, it is quite possible that nonconstant spreads will be found.

Until now, we have only used end-of-the-month data for December 1995. However, modeling the evolution of the term structure is an important problem in finance and is necessary, for example, to price interest rate derivatives (Jarrow 2002). To study this evolution, we fit the STRIPS and AT&T bond prices separately for each month over the 21-month period of April 1994 to December 1995. Figure 3 shows the evolution of the end-of-month forward rates estimated by P splines and EBBS, with a constant spread. If we fix maturity and observe the forward rate as a function of time, then we see a rough curve. This is to be expected, because interest rates move randomly and abruptly. This is why we did not use a bivariate smooth in both maturity and time.

APPENDIX: ASSUMPTIONS

The following assumption is needed for the proof of (strong) consistency.

Assumption A.1. The parameter space Θ is compact. The mean function m(·) is continuous on Θ, (1/n)Σ^n_i=1 (m_i(δ) - m_i(δ_0))^2 converges uniformly to some limit in δ, δ ∈ Θ, and Q(δ) = lim(1/n)Σ^n_i=1 [m_i(δ) - m_i(δ_0)]^2 exists and has a unique minimum at δ = δ_0.

Under the following additional assumption, asymptotic normality can be established.

Assumption A.2. The true parameter vector δ_0 is an interior point of Θ, the mean function m(·) is twice continuously differentiable in a neighborhood of δ_0, and δ_0 = lim(1/n)Σ^n_i=1 m_i^{(1)}(δ_0)m_i^{(1)}(δ_0) exists and is nonsingular, where m_i^{(1)}(δ_0) = ∂m_i/∂δ_j|δ_0. Furthermore, (1/n)Σ^n_i=1 m_i^{(2)}(δ)m_i^{(2)}(δ) exists and is nonsingular, where m_i^{(2)}(δ) = ∂^2m_i/∂δ_j∂δ_k|δ_0. Finally, (1/n)Σ^n_i=1 m_i^{(1)}(δ)m_i^{(1)}(δ) converges uniformly in δ to an open neighborhood of δ_0.

[Received December 2002. Revised October 2003]
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.